

Multimedia Appendix 2. Pseudo code of symptom frequencies based on vector distance (SF-DIST) and cosine similarity (SF-COS)

**SF-COS :**

**Input :** *patient\_symptoms (10x1), symptoms\_frequencies (10x4)*

if all *patient\_symptoms* == 0:  
    Return LOW RISK

For each column of *symptoms\_frequencies*:  
    *similarity* = *cosine\_similarity(patient\_symptoms, symptoms\_frequencies[column])*

*similarity* = *normalise(similarity)*

*similarity[COVID\_19]* = *similarity[COVID\_19]\*(Area\_Risk\_Factor + Contact\_Risk\_Factor)*

**Output :** *return disease with maximum similarity*

**SF-DIST :**

**Input :** *patient\_symptoms (10x1), symptoms\_frequencies (10x4)*

if all *patient\_symptoms* == 0:  
    Return LOW RISK

For each column of *symptoms\_frequencies*:  
    *similarity* = *sum(abs((patient\_symptoms - symptoms\_frequencies[column]))*

*similarity* = *normalise(similarity)*

*similarity[COVID\_19]* = *similarity[COVID\_19]\*(Area\_Risk\_Factor + Contact\_Risk\_Factor)*

**Output :** *return disease with maximum similarity*